

5th Annual Joint Bioinformatics Workshop

July 19, 2005
2229 Seamans Center
University of Iowa

by

Haitao Cheng
Iowa State University

Prediction of protein secondary structure by mining structural fragment database

ABSTRACT

A new method for predicting protein secondary structure from amino acid sequence has been developed. The method is based on multiple sequence alignment of the query sequence with all other sequences with known structure from the protein data bank (PDB) by using BLAST.

The fragments of the alignments belonging to proteins from the PDB are then used for further analysis. We have studied various schemes of assigning weights for matching segments and calculated normalized scores to predict one of the three secondary structures: α -helix, β -sheet, or coil. We applied several artificial intelligence techniques: decision trees (DT), neural networks (NN) and support vector machines (SVM) to improve the accuracy of predictions and found that SVM gave the best performance. Preliminary data show that combining the fragment mining approach with GOR V (Kloczkowski et al, Proteins 49 (2002) 154–166) for regions of low sequence similarity improves the prediction accuracy.